On the Utility of Mathematical Models and their Use in Evolutionary Biology

Yannis MICHALAKIS *, *Eric* WAJNBERG ** & *Carlos* BERNSTEIN ***

* Laboratoire d'Ecologie, URA 258 CNRS, Université P. & M. Curie, CC237,
7, quai Saint Bernard, 75252 Paris Cédex 05, France
** Centre de Recherche d'Antibes, Laboratoire de Biologie des Invertébrés,
Unité de Biologie des Populations, I.N.R.A., 37, Boulevard du Cap, 06600 Antibes, France
*** Biométrie, Génétique et Biologie des Populations, UMR 5558 CNRS, Université Lyon 1,
43, boulevard du 11 Novembre 1918, 69622 Villeurbanne Cédex, France

ABSTRACT

Mathematical modeling is a powerful tool used in almost any kind of scientific endeavor. Mathematical models are as diverse as the problems we tackle with them because different problems call for different methods. The recent spectacular development of computing capacity has dramatically increased the number of potential applications of mathematical models, since many more problems are now tractable numerically. Models can be characterized according to three essential properties: generality, realism and precision (LEVINS, 1966). These properties often trade-off and only the questions to which the model is supposed to give an answer, (*i.e.* evolutionary questions in the case of evolutionary biology), can guide the choice of the relevant degree of generality, precision or realism. Mathematical models can be a very useful tool in several respects. For instance, where intuition may be a poor guide, they can be used as guides to study qualitatively the behavior of very complex systems. They often help to clarify our ideas, and their results can be used to build null hypotheses. Some caution is needed, however, in their use, because it is often the case that several different models (or different combinations of parameter values of the same model) yield similar results compatible to biological observations. Considering alternative hypotheses is necessary and essential.

RÉSUMÉ

A propos de l'utilité des modèles mathématiques et de leur utilisation en biologie évolutive

La modélisation mathématique est un outil puissant utilisé quasiment dans tout champ d'activité scientifique. Il existe une grande diversité de modèles mathématiques, des problèmes différents nécessitant l'utilisation de méthodes différentes. Le récent et spectaculaire développement des capacités de calcul en informatique a augmenté radicalement le nombre de problèmes auxquels on peut essayer de répondre. Les modèles mathématiques ont trois propriétés essentielles ; généralité, réalisme et précision (LEVINS, 1966). En général il existe un antagonisme entre ces propriétés. Le degré de pertinence de chacune ne pourrait être guidé que par des considérations liées aux questions auxquelles le modèle est censé répondre. Les modèles mathématiques peuvent être utiles à plusieurs titres. Ils peuvent servir pour étudier le comportement qualitatif de systèmes complexes dans des cas où l'intuition est un guide peu fiable. Ils aident souvent à clarifier les idées, et leurs résultats peuvent constituer des hypothèses nulles. Cependant, ils doivent être utilisés avec précaution. En effet, il arrive souvent que des modèles qualitativement différents (ou bien des combinaisons différentes des valeurs de paramètres du même modèle)

produisent des résultats similaires et compatibles avec les observations biologiques. Ainsi, la prise en compte d'hypothèses alternatives est non seulement essentielle mais aussi nécessaire.

INTRODUCTION

Mathematical modeling is a very powerful tool used in almost any kind of human activity. The nature of mathematical models is highly diverse because different problems require different methods. The recent spectacular development of computing capacities and of software for mathematical computation has radically increased the number of potential applications of mathematical models, since many more problems are now tractable either by numerical calculations or by stochastic simulations.

In this paper, we first give some reasons on why to build models, we then outline some characteristics of models and finally discuss some applications of mathematical models, especially in population and evolutionary biology. Before proceeding any further we would like to warn the reader on several accounts. First, none of us was trained neither as a philosopher nor as a mathematician, but rather as a population biologist. This text should therefore be regarded only as a personnal account and opinion on what has become our main everyday activity. We, by no means, intend to represent the orthodoxy or establish rules on how or why mathematical models should be built or applied, nor do we seek to be exhaustive. Should the style of what follows seem dogmatic in some places, we would like to present our apologies to the reader in advance. Such authoritative-like expressions just reflect our wish to avoid overloading the text with continuous repetitions of the fact that this text reflects only our opinion. Second, we take it for granted that the reader is convinced of the utility of Theory in Science, actually of the impossibility to do Science without a theoretical framework. Finally, citations to models or theories mentioned here should be solely regarded as illustrations of a particular point we want to make, and should by no means be viewed as a judgement, be it positive or negative, of their scientific merit.

WHY BUILD/USE MODELS?

There are several good reasons to build mathematical models. We will concentrate on the three that we think are most important in evolutionary biology. Usually, models are built in order to test the effect of some changes in biological processes which can be evaluated this way. In this respect, models often allow a conceptual synthesis. A very good example is the development of theoretical population genetics in the first half of the 20th century by FISHER, HALDANE and WRIGHT (PROVINE, 1971). These three scientists, by developing mathematical models, managed to make the large majority of biologists accept DARWIN's theory of evolution, by showing that natural selection of random mutations was indeed able to account for observed patterns of adaptation. What is really remarkable is that the development of this impressive body of theory took place while most of the underlying mechanisms concerning the way genes function or even what genes are, were largely unknown.

While allowing a conceptual synthesis, mathematical models very often generate new hypotheses. Such hypotheses can be subsequently theoretically or experimentally evaluated. It is very often the case that the potential role of a particular mechanism has been ignored until

revealed by a modeling exercise. In other cases the relevance of a given mechanism as an explanation of a particular set of observations cannot be easily assessed experimentally. A model can suggest simpler ways to test this mechanism or at least help to explore its relevance.

Finally mathematical models, in that they are often quantitatively predictive, can be very useful as aids to decision making and constitute null hypotheses against which predictions can be tested. For instance, the development of the neutral theory of molecular evolution would have been impossible without the mathematical models developed by Motoo KIMURA and his colleagues (KIMURA, 1983). The mathematical models, in conjunction with statistical tests based on these models and developed by others, not only substantiated the plausibility of this theory, but provided a null hypothesis against which observations of molecular polymorphism are compared.

PROPERTIES AND KINDS OF MODELS

It is possible to classify models with respect to three properties: generality, realism and precision. As LEVINS (1966) noted, there is usually a sort of trade-off between these properties. For example, to be qualified as "precise" a model would have to take into account many detailed conditions and processes which apply to the situation modeled, most probably to the expense of other processes which do not apply to the modeled situation. By doing so, the model would very likely be quite "realistic" as well, but it is highly unlikely that it would be "general" in the sense that it will probably not be applicable to a different situation without modification.
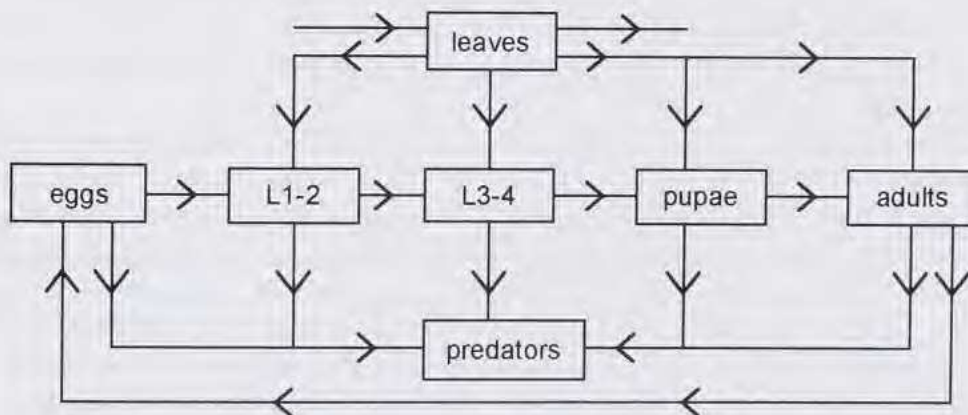

FIG. 1. — Schematic representation of a tritrophic food chain involving a herbivorous insect (represented by its various developmental stages), its host plant (represented by its leaves) and its predators.

The optimal mixture of these three properties can be defined only by the question addressed and the generality that one wishes the answer to have. A given model will be very precise, perhaps too precise, to answer a given question and too general to answer another one. To illustrate this point, consider the model depicted in figure 1: this model represents a tritrophic system consisting of host plants, herbivorous insects feeding on these plants, and predators feeding on the insects. This model could be considered adequate to study the evolution of

herbivorous insect populations. It would be too detailed, however, if one wanted to address questions on tritrophic systems in general, since many tritrophic systems involve organisms which, for instance, do not have as many distinct developmental stages as insects. On the other hand this model is probably too general for other questions one might want to address on herbivorous insect tritrophic systems. For example, plants are only represented by the amount of leaves produced, assuming that all leaves are equivalent. This is probably an oversimplification as quality of leaves of a given plant typically changes with leave age, while the quality of leaves of different plants can be very different. Exactly the same arguments apply to the predators.

STEPS IN MODEL BUILDING

The most important step in mathematical modeling in evolutionary biology is the very beginning: identify an interesting biological question. If the biological question is not relevant, then, obviously, anything that will follow will be equally irrelevant from a biological perspective (though may still be relevant from an applied mathematics perspective). Once the relevant biological question has been identified, one has to think of the different processes that might contribute to generate the observed patterns. The next step would be the formalization in mathematical terms of the interactions of the various processes identified in the previous step. This implies important decisions about which are the appropriate simplifications to be made. Then, the mathematical model can be analyzed, analytically or by computer simulations, and the results interpreted and discussed. Quite often the result of this analysis and discussion generates new hypotheses, most frequently due to the examination of further observations and the emergence of processes that might be involved and which had not be taken into account so far. The activity of model building resembles to the construction of a spiral, where one starts with an initial idea which is reexamined in the light of the mathematical analysis and then expands to another idea by the incorporation of new elements.

A step often discussed is the "validation" of a model. What is at stakes is not whether a mathematical model is mathematically correct (such cases are "relatively" easily resolved), but rather how does it apply to the biological question it is supposed to answer. Such situations may arise for several reasons. First, some relevant mechanisms may not be taken into account by the specific model. In this case, the model as such cannot answer the question it was supposed to, but requires, at least, modifications. We would like to express at this point the view that such "failures" are equally valuable for the process of scientific knowledge as "negative" experiments, *i.e.* experiments which fail to find statistically significant effects of the factors they have examined, and should therefore deserve equal attention. A second reason for which a model may not answer a biological question is because some specific parameter values used to produce quantitative predictions with the model are wrong. In this case, the problem is not actually due to the model but to its applications. This kind of problem can be "easily", at least from the model's perspective, solved by implementing the model with new parameter values and does not require the modification of the model itself.

In practice, of course, things are much more complicated than what would appear from the previous argumentation. The first, and probably most important problem, is how one decides that a given model does not satisfactorily explain the observed patterns. The answer to this problem is

clearly question-specific. If the question requires a quantitative answer one can conduct experiments and actually empirically measure the variables whose behavior the model is designed to predict. Because quantitative variables can be easily accompanied by confidence intervals one may compare the values predicted by the model using the observed confidence intervals. For questions requiring qualitative answers, however, it is much more difficult to decide whether a specific model provides a satisfying answer. The main difficulty lies in the fact that the way the fit of the model is ascribed as "good" or "bad" remains largely subjective. For instance, would a model designed to explain the evolution of dioecy in Angiosperms and which actually explains about 70% of the data be considered as validated or invalidated? Furthermore, the fact that models are not a single hypothesis but rather a full set of them makes it difficult to decide whether to just reject an unsatisfactory model or modify it.

Another problem arises because many models may lead to the same patterns, at least for a particular variable. Therefore the fact that a particular pattern can be satisfactorily explained by a given model should not be viewed as a confirmation of the model, but rather as the absence of an information. The only way to avoid such caveats is to examine the alternative models and force them to produce contradicting predictions with respect to at least one experimentally measurable variable. In saying this, we do not want to imply that scientists should not publish papers with mathematical models unless they have examined all alternative models. We just suggest that until such an examination is done by somebody the question should still be considered open.

CONCLUDING REMARKS

Mathematical modeling is a very powerful tool, which is necessary in almost any research program. Models are necessary because they allow to build quantitative null hypotheses. As evolutionary biology progresses towards the study of problems of increasing complexity the use of mathematical models becomes indispensable, simply because the increasing complexity of the questions addressed renders intuitive expectations untrustly and in some cases even impossible.

Mathematical modeling offers a relatively inexpensive way to contrast alternative hypotheses and thus allow the evaluation of the relative importance of several mechanisms. For such a parallel examination to be possible, alternative hypotheses should be forced to produce opposite predictions. The interested reader can find more information on mathematical modeling applied in biology in BROWN & ROTHERY (1993) and in PROVINE (1971).

"Models" are used in many areas of human activities though they do not always bear the same meaning. In the arts, a "model" is "what really exists", what the Artist wants to represent by her/his work. In biology a "model" is meant to be a simplified representation of "what really exists". Such representations will only fulfill their role if they are guided by natural history, so that the final object of investigation remains clear in the mind of all actors.

REFERENCES

BROWN, D. & ROTHERY P., 1993. — *Models in Biology: Mathematics, Statistics and Computing*. Chichester, John Wiley & Sons: 1-708

KIMURA, M., 1983. — *The Neutral Theory of Molecular Evolution*. Cambridge, Cambridge University Press: 1-367

LEVINS, R., 1966. — The strategy of model building in population biology. *The American Scientist*, **54**: 421-431.
MURRAY, J. D., 1993. — *Mathematical Biology*. Berlin, Springer Verlag, 2nd edition: 1-768
PROVINE, W. B., 1971. — *The Origins of Theoretical Population Genetics*. Chicago, Chicago University Press: 1-202